



The transmission process: A combinatorial stochastic process for the evolution of transmission trees over networks



Raazesh Sainudiin ^{a,*}, David Welch ^b

^a Laboratory for Mathematical Statistical Experiments, Christchurch Centre and Biomathematics Research Centre, School of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch 8041, New Zealand

^b Computational Evolution Group and Department of Computer Science, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

HIGHLIGHTS

- We derive a combinatorial stochastic process for the evolution of the transmission tree over the host contact network in a susceptible–infected (SI) epidemic model.
- We develop a biparametric Beta-splitting model that generates transmission trees without explicitly modelling the underlying contact network.
- We show that for specific values of the parameters we can get the exact probabilities for a complete, star or path network.
- We use the maximum likelihood estimator to consistently infer the two parameters driving the transmission process based on observations of the transmission trees.

ARTICLE INFO

Article history:

Received 19 January 2016

Received in revised form

22 July 2016

Accepted 22 July 2016

Available online 9 August 2016

MSC Classification:

92D30

05C05

05C20

60J10

60J27

05C85

Keywords:

Rooted ranked planar binary tree

Contact network

Susceptible–infected epidemic model

Non-parametric combinatorial stochastic process

Parametric Beta-splitting model

Random graph models

Meme evolution

ABSTRACT

We derive a combinatorial stochastic process for the evolution of the transmission tree over the infected vertices of a host contact network in a susceptible–infected (SI) model of an epidemic. Models of transmission trees are crucial to understanding the evolution of pathogen populations. We provide an explicit description of the transmission process on the product state space of (rooted planar ranked labelled) binary transmission trees and labelled host contact networks with SI-tags as a discrete-state continuous-time Markov chain. We give the exact probability of any transmission tree when the host contact network is a complete, star or path network – three illustrative examples. We then develop a biparametric Beta-splitting model that directly generates transmission trees with exact probabilities as a function of the model parameters, but without explicitly modelling the underlying contact network, and show that for specific values of the parameters we can recover the exact probabilities for our three example networks through the Markov chain construction that explicitly models the underlying contact network. We use the maximum likelihood estimator (MLE) to consistently infer the two parameters driving the transmission process based on observations of the transmission trees and use the exact MLE to characterize equivalence classes over the space of contact networks with a single initial infection. An exploratory simulation study of the MLEs from transmission trees sampled from three other deterministic and four random families of classical contact networks is conducted to shed light on the relation between the MLEs of these families with some implications for statistical inference along with pointers to further extensions of our models. The insights developed here are also applicable to the simplest models of “meme” evolution in online social media networks through transmission events that can be distilled from observable actions such as “likes”, “mentions”, “retweets” and “+1s” along with any concomitant comments.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The detailed picture of the path an epidemic takes through a population over its course is encapsulated in the *transmission tree*. The transmission tree represents the physical continuum of contacting hosts and thus frames the host-level structure within which pathogens are transmitted in a communicable disease.

* Corresponding author. Current address: Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden.

E-mail addresses: raazesh.sainudiin@gmail.com (R. Sainudiin), david.welch@auckland.ac.nz (D. Welch).

Therefore, models of transmission trees are crucial to understanding the evolution of pathogen populations. Constructing models of transmission trees is the main focus of this paper. Although we limit ourselves here to the epidemiological context of transmissions of a communicable disease over a contact network of hosts for concreteness of language and notions from a field with a longer research history, most of our basic results and insights are naturally applicable, as briefly discussed in Section 5.2, to the cultural context of transmissions of “memes” (Dawkins, 1976 p. 192) over a social network of individuals, such as Twitter (Solon, 2013). More generally, they can be used to model transmission events in Finite Markov Information Exchange processes (Aldous, 2013, Section 2.2) as described below.

To understand the process by which a transmission tree grows, we need to consider (i) the *structure of the population* in which the epidemic spreads and (ii) the *state of the individuals* in the population as the epidemic spreads. Network models are a natural candidate for describing population structure where the population is identified with a network in which each vertex represents an individual and an arc (a directed weighted edge) from vertex i_i to i_j , given by a non-negative $w_{i,j} \in [0, \infty)$, represents the propensity with which the infection can be transmitted from i_i to i_j . This propensity can be given meaning in terms of *frequency of contacts* by taking each $w_{i,j} > 0$ to specify independent rate- $w_{i,j}$ Poisson process for the contact times between i_i and i_j , for instance (this is the “meeting process” of Aldous, 2013). We call these networks *contact networks* and assume that they are fixed or static through time. Thus, the contact network of a population summarizes “who can contact whom and how frequently” and is depicted in Fig. 1(a) for a small population with vertices labelled by individuals i_1, i_2, \dots, i_9 (the edges are undirected). Note that we sometimes label the vertices starting from i_0 to stay true to the indexing convention in sageMath/python (but this should be clear from the context).

The epidemic state of each individual at a given time can be in one of the several possible states, depending on the particularities of the epidemic model. The simplest case, known as the SI model, involves only two states that indicate whether an individual at a given time is susceptible (S) to or infected (I) by a pathogen. Under this model, the only possible state transition is from S to I as specified by the contact network. In other words, a susceptible individual can be infected by any individual in its in-neighborhood who is already infected. The contact network with its individual vertices further “tagged” by their epidemic states (S or I) is called the *tagged contact network*. The epidemic states of the individuals in the population after some time are shown by tagging (coloring) the infected or susceptible individuals with I or S tags (red or white colors) in Fig. 1(b).

The *transmission digraph* is a directed edge-labelled subgraph of the contact network containing all infected vertices and directed edges labelled by the time of transmission. It is a basic object of interest and is depicted in Fig. 1(b). The transmission digraph can also be represented by the more convenient *transmission tree* shown in Fig. 1(c). The internal vertices of the transmission tree correspond to times of transmission events, the below (or left) and above (or right) planar sub-trees encode who infected whom, and

the leaf vertices correspond to the set of infected individuals. Since the tagged contact network co-evolves with the transmission tree, the transmission process is naturally seen as a Markov chain on the product space of tagged contact networks and transmission trees. We consider a stochastic model, as opposed to a deterministic one, to be natural because the spread of an epidemic is inherently probabilistic (Andersson and Britton, 2000).

The transmission tree captures several details about how an infection spreads through the population, including combinatorial structural information such as who infected whom, order and timing of infection events, the time it takes for a specified set of individuals to be infected, tree shape statistics such as indices of Sackin (1975) and Colless (1982), number of cherries or sub-terminal vertices (McKenzie and Steel, 2000), etc., various isomorphism classes, such as (un)ranked/(non)planar unlabelled trees and so on, but also classical epidemiological univariate statistics, such as prevalence and incidence through time, reproduction numbers and total time of epidemic.

Furthermore, by a natural extension of the pure-birth process underpinning the SI model to a birth-and-death process that is combinatorially more involved with an additional epidemic state indicating whether the individual is “removed” (R) from the population, one can extend the transmission process developed here for the SI epidemic model to the more realistic susceptible–infected–recovered (SIR) epidemic model. With such an extension, which we will not pursue in this elementary study of the simplest SI epidemic model (for reasons explained below), the leaves of the SIR transmission trees will not only be tagged by I but also by R and they will naturally capture various univariate statistics of interest to applied epidemiologists including the final-size or total number of infections (Ludwig, 1975; Pellis et al., 2008; House et al., 2012). We outline a set of combinatorial steps needed towards such a future direction of work in Section 5.1.

While various analytical results (e.g. Andersson and Britton, 2000) and computationally intensive methods (e.g. House et al., 2012) are available for various univariate epidemiological statistics and can often be obtained without explicitly modelling the tree, most insights about the structural information in the tree (even for the simplest SI epidemic model) are difficult to derive analytically and so are based on simulation studies over parametric families of specific models.

Empirical efforts to understand the transmission process have historically focused on time series and individual event times (such as infection or recovery times) as the main data source. These relatively sparse forms of data have been difficult to collect and not particularly informative, providing limited information about the transmission tree (but see Haydon et al., 2003; Wallinga and Teunis, 2004) or the underlying contact network.

Recently, there has been an increasing attention paid to using the large amounts of viral and bacterial genomic data now available to study outbreaks. The key observation suggesting this data will be informative about the transmission tree is that, if there is little within-host viral genetic diversity, the phylogenetic tree of pathogenic genomes will match the transmission tree (though, in many cases, this assumption does not hold, Romero-Severson et al., 2014; Ypma et al.,

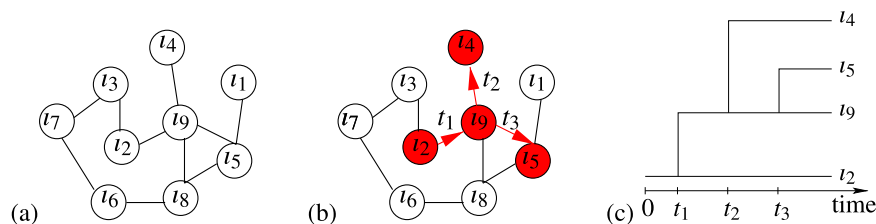


Fig. 1. Spread of an epidemic over (a) the contact network of a population as shown by (b) a sub-network where edges representing transmission events are labelled by the time of event and the infected vertices are colored red and (c) the corresponding transmission tree. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Download English Version:

<https://daneshyari.com/en/article/6368853>

Download Persian Version:

<https://daneshyari.com/article/6368853>

[Daneshyari.com](https://daneshyari.com)